Lab XX – MATH 240 – Computational Statistics

First Author
Affiliation
Department
email@domain

Abstract

First part of a three-part lab that involves working with the program Essentia in order to retrieve data from song libraries in order to determine how much a particular artist has contributed to a work with multiple authors. As part of task 1, I created a batch file that automatically processes the Essentia command for every song within the MUSIC directory. Following this, I became familiar with how to use the jsonlite package for R, and extracted data using this package from an example song that was provided within my assignment's repository.

Keywords: For Loops; Packages; Lists;

1 Introduction

Week 2's lab assignment was the first part of a three-part project that involves the sourcing and extraction of music data from a program called Essentia in an effort to determine the degree of contribution among different bands collaborating on the same song. This program allows us to extract a wide variety of data from songs, though we must use its command prompt to access it. However, Essentia's command prompt must be entered for every single song within the directory that you are looking to extract data from, which can take quite some time depending on the size of the song directory. To remedy such, task one of the assignment involves creating a batch file that will automate the generation of commands for all songs which can be used to process the entire song directory at once. After creating our batch file within part one, we begin proccessing json files through the jsonlite package for R, which allows us to extract the data for each song from Essentia.

2 Methods

As mentioned above, part one of this assignment involves the creation of a batch file in order to process every single song within our music directory through Essentia simultaneously. To create this batch file, a vector was created containing the names of all files in our music directory, so that we'd be able to reference them within R. Then, we used the strcount function from the stringr package in order to subset all directories within the MUSIC folder into the classifiation of artist, album, and track. Following such, two loops were established in order to iterate among all albums and song within the MUSIC folder. Within these loops, the strcount function was used to extract the name of the artist, album, and track from each song with each being assigned to an object. These objects were then pasted together with our necessary command line, and saved to an empty vector that would populate with every iteration of the loop. This code was then written to a text file, ready to be executed as a batch. For the second part of the lab, we processed an example json file using the jsonlite package for R. Using the fromJSON function, I was able to extract all Essentia data from the provided example json file. Furthermore, specific values, including: average loudness, mean of spectral energy, dancability, bpm, key, scale, and length were extracted using commands found on Essentia's website. Each type of data was then assigned to objects and printed using the paste function.

3 Results

Tie together the Introduction – where you introduce the problem at hand – and the methods – what you propose to do to answer the question. Present your data, the results of your analyses, and how each reported aspect contributes to answering the question. This section should include table(s), statistic(s), and graphical displays. Make sure to put the results in a sensible order and that each result contributes a logical and developed solution. It should not just be a list. Avoid being repetitive.

3.1 Results Subsection

Subsections can be helpful for the Results section, too. This can be particularly helpful if you have different questions to answer.

Bibliography: Note that when you add citations to your bib.bib file *and* you cite them in your document, the bibliography section will automatically populate here.

4 Appendix

If you have anything extra, you can add it here in the appendix. This can include images or tables that don't work well in the two-page setup, code snippets you might want to share, etc.